4.1.4 Mixed Linear Effects Models

A fixed general mean μ or a fixed intercept term b_0 and random residual term e occur in almost all models that were presented so far. Apart from these, all other effects were either all fixed or random[1]. We now consider models where some effects (other than μ and e) are fixed and some are random. Such models are called **mixed linear effects models**[2].

An example dataset which could be analysed with a mixed linear effects model would be, if we would add to each animal in our reference dataset on body weight, breast circumference and breed also the sire of each animal. If some of these animals would share the same sire and hence would be half sibs, the dataset would again as already seen in the repeated observations data, a specific variance structure. This is due to the fact that body weights from half sibs would be expected to be more similar than observations from unrelated animals.

Table 4.5: Body Weight, Breast Circumference, Breed and Sire of Beef Cattle Animals

Animal	Body Weight	Breast Circumference	Breed	Sire
1	471	176	Angus	S1
2	463	177	Angus	S1
3	481	178	Simmental	S3
4	470	179	Angus	S2
5	496	179	Simmental	S3
6	491	180	Simmental	S4
7	518	181	Limousin	S5
8	511	182	Limousin	S5
9	510	183	Limousin	S6
10	541	184	Limousin	S6

When fitting a mixed linear effects model to a dataset as shown in Table 4.5, the question is which effects should be taken as fixed and which should be considered to be random. As already mentioned in this case, `Breast Circumference` and `Breed` would be modelled as fixed effects and `Sire` would be modelled as a random effect. In general, there are not strict rules that would tell us which effects should be modelled as fixed effects an which ones should be considered as random. In our dataset we can certainly say that for `Breast Circumference` and `Breed` we are interested in the effect sizes of the values that are observed in the given datasets. In contrasts to that, we can say that the included sires

[1] Except for a small introduction into repeated measures models, we have not really look at random models in great detail. But they are not of great importance to the treatment of mixed models.

[2] Sometimes these models are just called mixed models. We are using these terms interchangably

are a random sample of a larger population of sires. Furthermore, the primary interest in the sire effects are in the imposed covariance structure of the data due to the sire effects. In the case where the primary interest is in the variance imposed by a certain effect, then the respective effect has to be modelled as a random effect.

The general mixed effects model can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{4.9}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{u}$ is the vector of random effects, $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices and $\mathbf{e}$ is the vector of random residuals. The random effects are assumed to have expected values of zero and given specific variance-covariance matrices. Hence we can write

$$E \left[\begin{array}{c} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{array}\right] = \left[\begin{array}{c} \mathbf{X}\mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \end{array}\right] \tag{4.10}$$

The variance-covariance matrices are specified as

$$var \left[\begin{array}{c} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{array}\right] = \left[\begin{array}{ccc} \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathbf{T}} + \mathbf{R} & \mathbf{Z}\mathbf{D} & \mathbf{R} \\ \mathbf{D}\mathbf{Z}^{\mathbf{T}} & \mathbf{D} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{array}\right] \tag{4.11}$$

with $var(\mathbf{u}) = E(\mathbf{u}\mathbf{u}^{\mathbf{T}}) = \mathbf{D}$ and $var(\mathbf{e}) = E(\mathbf{e}\mathbf{e}^{\mathbf{T}}) = R$.

Assuming $\mathbf{V}$ is not singular, the normal equations stemming from the generalized least squares are

$$\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\mathbf{b}^0 = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{4.12}$$

with a solution

$$\mathbf{b}^0 = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y} \tag{4.13}$$

From that solution, we can get estimates of estimable functions for the fixed effects as previously discussed for fixed models.

For the random effects $\mathbf{u}$, the conditional expectation of $\mathbf{u}$ given the observations $\mathbf{y}$ are of particular interest as estimators. Assuming multivariate normality for $\mathbf{u}$ and $\mathbf{e}$, we can write

$$\begin{aligned} \hat{\mathbf{u}} = E(\mathbf{u}|\mathbf{y}) &= E(\mathbf{u}) + cov(\mathbf{u}, \mathbf{y}^T)(var(\mathbf{y}))^{-1}(\mathbf{y} - E(\mathbf{y})) \\ &= \mathbf{D}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) \end{aligned} \tag{4.14}$$

Both terms, the solution for $\mathbf{b}^0$ and the estimate $\hat{\mathbf{u}}$ depend on the inverse matrix $\mathbf{V}^{-1}$ which can be extremely large and difficult to compute. In different publications, the research group of Charles Henderson has shown that solving the following system of equations leads to the same estimates for both the fixed and the random effects. This system of equations is called **Mixed Model Equations** and is shown below.

$$\left[\begin{array}{cc} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{array} \right] \left[\begin{array}{c} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{array} \right] = \left[\begin{array}{c} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{array} \right] \tag{4.15}$$

4.2 Pedigree BLUP

The linear mixed effects models as shown above can be applied to datasets in livestock breeding. In such a model, the response variable y corresponds to measurements or observations of phenotypic traits. The vector of fixed effects b contains all information about the known environment such as `Breed`, `Herd`, `Season`, `Age` and possibly other predictors that have an influence on the response. The random effects u contain the breeding values of animals of interest in our livestock breeding population. Once all the informations of the data are collected, it can be transfered into model components. The model components are then used to construct the mixed model equations. Solutions to these equations provide estimates of fixed effects and predictions of breeding values. Properties of the predicted breeding values can be summarized as

- Best: the predicted breeding values have minimum prediction error variance
- Linear: the predicted breeding values are linear functions of the data
- Unbiased: the expected value of the predicted breeding values is equal to the expected value of the true breeding value
- Prediction: because breeding values cannot be observed, the results are called predictions.

The above listed properties are often abbreviated as BLUP.

The application of linear mixed effects models to livestock breeding datasets can be done in two different ways.

1. Sire model: only sires in the dataset get breeding values
2. Animal model: all animals in a datasets (also parents without observations) get breeding values

4.2.1 Sire Model

In a sire model the vector $\mathbf{u}$ of random effects contains all sires in the dataset. For the example data shown in Table 4.5, this corresponds to

$$\mathbf{u} = \begin{bmatrix} S1 \\ S1 \\ S3 \\ S2 \\ S3 \\ S4 \\ S5 \\ S5 \\ S6 \\ S6 \end{bmatrix}$$

Because the sire breeding values ($\mathbf{u}$) are random effects, we also have to specify the expected value and the variance-covariance matrix of $\mathbf{u}$. Because breeding values are defined as deviations, the expected values of the sire breeding values are zero. Hence

$$E(\mathbf{u}) = \mathbf{0} \tag{4.16}$$

$$var(\mathbf{u}) = \mathbf{D} \tag{4.17}$$

with $\mathbf{D}$ beeing the variance-covariance matrix between the sire breeding values. If the sires are not related, then $\mathbf{D} = \sigma_s^2\ I$ where σ_s^2 is a sire variance component. If the sires are related then $\mathbf{D} = \sigma_s^2\ \mathbf{A}_s$ where $\mathbf{A}_s$ is the sire relationship matrix containing elements of probabilities of sharing allels based on identity by descent between related sires as off-diagonal elements. The diagonal elements of $\mathbf{A}_s$ are all one.

For the moment, we assume that the variance component such as σ_s^2 are all given. In reality, such components would also need to be estimated from the data. The discussion on how to estimate variance components from the data is deferred to a later chapter.

4.2.2 Animal Model

The major difference between the sire model and the animal model is that in the animal model all animals in the dataset receive breeding values. Hence in the dataset shown in Table 4.5, we would need to add the dams.

Table 4.6: Body Weight, Breast Circumference, Breed, Sire and Dam of Beef Cattle Animals

Animal	Body Weight	Breast Circumference	Breed	Sire	Dam
1	471	176	Angus	S1	D1
2	463	177	Angus	S1	D2
3	481	178	Simmental	S3	D3
4	470	179	Angus	S2	D2
5	496	179	Simmental	S3	D3
6	491	180	Simmental	S4	D4
7	518	181	Limousin	S5	D5
8	511	182	Limousin	S5	D5
9	510	183	Limousin	S6	D6
10	541	184	Limousin	S6	D7

The vector **u** contains breeding values for all animals in the dataset, also from parents that do not have observations. Hence

$$\mathbf{u} = \left[\begin{array}{c} S1 \\ S2 \\ ... \\ D1 \\ D2 \\ ... \\ 1 \\ 2 \\ ... \\ 10 \end{array}\right]$$

The expected value and the variance-covariance matrix of **u** are defined as

$$E(\mathbf{u}) = \mathbf{0} \tag{4.18}$$

$$var(\mathbf{u}) = \mathbf{D} = \mathbf{A}\sigma_u^2 \tag{4.19}$$

where the matrix **A** corresponds to the numerator relationship matrix. This matrix contains the probabilities of two animals sharing alleles identical by descent on the offdiagonal elements. The diagonal elements of **A** are computed as one plus the inbreeding coefficient of an animal. The inbreeding coefficient of an animal is given by half of the relationship coefficient of the parents.